

# $\alpha$ IIb $\beta$ 3 variants defined by next-generation sequencing: Predicting variants likely to cause Glanzmann thrombasthenia

Lorena Buitrago<sup>a,1</sup>, Augusto Rendon<sup>b,c,d,1,2</sup>, Yupu Liang<sup>e,1</sup>, Ilenia Simeoni<sup>b,d</sup>, Ana Negri<sup>f</sup>, ThromboGenomics Consortium<sup>b,d,3</sup>, Marta Filizola<sup>f</sup>, Willem H. Ouwehand<sup>b,d,g,4</sup>, and Barry S. Coller<sup>a,4,5</sup>

<sup>a</sup>Allen and Frances Adler Laboratory of Blood and Vascular Biology and <sup>e</sup>Research Bioinformatics, The Rockefeller University, New York, NY 10065; <sup>b</sup>Department of Haematology, University of Cambridge, Cambridge CB2 0PT, United Kingdom; <sup>c</sup>Medical Research Council Biostatistics Unit, Cambridge Biomedical Campus, Cambridge Institute of Public Health, Cambridge, United Kingdom; <sup>d</sup>National Health Service Blood & Transplant, Cambridge, United Kingdom; <sup>f</sup>Department of Structural and Chemical Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029; and <sup>g</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

Edited by Dennis A. Carson, University of California, San Diego, La Jolla, CA, and approved March 5, 2015 (received for review November 24, 2014)

Next-generation sequencing is transforming our understanding of human genetic variation but assessing the functional impact of novel variants presents challenges. We analyzed missense variants in the integrin  $\alpha$ IIb $\beta$ 3 receptor subunit genes *ITGA2B* and *ITGB3* identified by whole-exome or -genome sequencing in the ThromboGenomics project, comprising ~32,000 alleles from 16,108 individuals. We analyzed the results in comparison with 111 missense variants in these genes previously reported as being associated with Glanzmann thrombasthenia (GT), 20 associated with alloimmune thrombocytopenia, and 5 associated with aniso/macrothrombocytopenia. We identified 114 novel missense variants in *ITGA2B* (affecting ~11% of the amino acids) and 68 novel missense variants in *ITGB3* (affecting ~9% of the amino acids). Of the variants, 96% had minor allele frequencies (MAF) < 0.1%, indicating their rarity. Based on sequence conservation, MAF, and location on a complete model of  $\alpha$ IIb $\beta$ 3, we selected three novel variants that affect amino acids previously associated with GT for expression in HEK293 cells.  $\alpha$ IIb P176H and  $\beta$ 3 C547G severely reduced  $\alpha$ IIb $\beta$ 3 expression, whereas  $\alpha$ IIb P943A partially reduced  $\alpha$ IIb $\beta$ 3 expression and had no effect on fibrinogen binding. We used receiver operating characteristic curves of combined annotation-dependent depletion, Polyphen 2-HDIV, and sorting intolerant from tolerant to estimate the percentage of novel variants likely to be deleterious. At optimal cut-off values, which had 69–98% sensitivity in detecting GT mutations, between 27% and 71% of the novel  $\alpha$ IIb or  $\beta$ 3 missense variants were predicted to be deleterious. Our data have implications for understanding the evolutionary pressure on  $\alpha$ IIb $\beta$ 3 and highlight the challenges in predicting the clinical significance of novel missense variants.

Glanzmann | integrin | single-nucleotide variants | next-generation sequencing | molecular modeling

Next-generation sequencing is transforming our understanding of human genetic variation (1) and providing profound insights into the impact of both inherited and de novo variants on human health (2, 3). At the same time, the data from these studies present serious challenges in providing information to individuals who are found to have variant forms of different proteins. To highlight these challenges, in this report we describe our experience in analyzing missense variants of the platelet  $\alpha$ IIb $\beta$ 3 integrin receptor from The Human Genome Mutation Database (HGMD), the 1000 Genomes project (1000G), the United Kingdom 10K Whole Exome Sequencing project (U.K.10KWES), the United Kingdom 10K Whole Genome Sequencing project (U.K.10KWGS), and The National Heart, Lung and Blood Institute Exome Sequencing Project (ESP); the latter four sources encompass ~32,000 alleles derived from 16,108 individuals.

The  $\alpha$ IIb $\beta$ 3 receptor has a number of virtues as a model system. First, it is required for hemostasis because platelet aggregation requires cross-linking of the activated form of  $\alpha$ IIb $\beta$ 3 by macro-

molecular ligands (4). Thus, defects in its biogenesis, activation, or ligand binding lead to the rare bleeding diathesis Glanzmann thrombasthenia (GT), an autosomal recessive disorder (5, 6). Patients with GT come to medical attention because of their hemorrhagic symptoms, and thus have been carefully analyzed clinically and with tests of platelet function for nearly 50 y (5, 7). The biochemical and molecular abnormalities in GT have been studied for nearly 40 y (4, 6, 8–10). In the past 10 y, high-resolution crystallography, electron microscopy, and computational studies of the  $\alpha$ IIb $\beta$ 3 receptor have provided atomic-level information on the correlation between receptor structure and function (11–21). In addition, ethnic groups with relatively high prevalence of GT have been defined that share the same genetic abnormality based on founder mutations, and the dates that some of the mutations entered the population have been estimated (22–28). An on-line registry of GT abnormalities, including patient phenotypes was developed in 1997 (29) and currently contains 51  $\alpha$ IIb and 43  $\beta$ 3 missense variants linked to

## Significance

Next-generation sequencing is identifying millions of novel gene variants, presenting challenges to researchers and clinicians. Variations in the genes *ITGA2B* and *ITGB3* affect integrin  $\alpha$ IIb $\beta$ 3, leading to the bleeding disorder Glanzmann thrombasthenia. We analyzed novel missense variants on ~32,000 alleles of *ITGA2B* and *ITGB3* and found missense variants affecting ~10% of the amino acids in each protein in ~1.3% of the population. Almost all variants are rare, indicating recent entry into the population. Two novel variants we predicted would be deleterious profoundly affected recombinant protein expression. At cut-off values that correctly predicted at least 69% of the known Glanzmann thrombasthenia mutations as deleterious, three variant prediction algorithms predicted that at least 27% of the novel variants are deleterious.

Author contributions: L.B., A.R., M.F., W.H.O., and B.S.C. designed research; L.B., A.R., Y.L., I.S., A.N., and M.F. performed research; ThromboGenomics Consortium contributed new reagents/analytic tools; L.B., A.R., Y.L., I.S., A.N., M.F., W.H.O., and B.S.C. analyzed data; and L.B., A.R., Y.L., M.F., W.H.O., and B.S.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

<sup>1</sup>L.B., A.R., and Y.L. contributed equally to this work.

<sup>2</sup>Present address: Genomics England Ltd., London EC1M 6BQ, United Kingdom.

<sup>3</sup>A complete list of the ThromboGenomics Consortium can be found in the *SI Methods*.

<sup>4</sup>W.H.O. and B.S.C. contributed equally to this work.

<sup>5</sup>To whom correspondence should be addressed. Email: collerb@rockefeller.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1422238112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1422238112/-DCSupplemental).

the disorder ([sinaicentral.mssm.edu/intranet/research/glanzmann](http://sinaicentral.mssm.edu/intranet/research/glanzmann)). The frequency of GT in the general population has not been established but it has a world-wide distribution, and based on data from hematologic practices, it is rare except in areas with a high rate of consanguineous mating (30).

Second, alloimmune disorders, including neonatal thrombocytopenia and posttransfusion purpura, due to amino acid substitutions in either  $\alpha$ Ib or  $\beta$ 3, have been characterized at the molecular biological level and correlated with mechanisms of immunologic recognition (31).

Third, inherited macrothrombocytopenia and anisothrombocytopenia have been associated with heterozygous missense variants or deletions in  $\alpha$ Ib or  $\beta$ 3. All of these appear to induce constitutive activation of the receptor and impair proplatelet formation (32–38).

Fourth,  $\alpha$ Ib $\beta$ 3 contributes to pathological platelet thrombus formation in human ischemic cardiovascular disease and  $\alpha$ Ib $\beta$ 3 is a validated target for antithrombotic therapy (39–41).

Fifth,  $\alpha$ Ib $\beta$ 3 is a member of the large integrin family of receptors, which includes 24 receptors derived from 18  $\alpha$ - and 8  $\beta$ -subunits (41, 42). These receptors are involved in important biologic processes, including development, cell migration, homing, cell survival, and adaptive immunity (41–43). More is known about the structure–function relationships of  $\alpha$ Ib $\beta$ 3 than the other members of the group, and so it serves as the paradigmatic integrin receptor (44, 45).

Sixth, 3D molecular models have been built based on crystallographic and NMR data to analyze the effects of novel amino acid substitutions on receptor structure and function and the generation of alloantigens (15, 46–53). The data from these models and assessments of the severity of the amino acid change in the variants have the potential to aid in predicting whether a novel variant is likely to affect receptor function and immunogenicity (54–59).

## Results

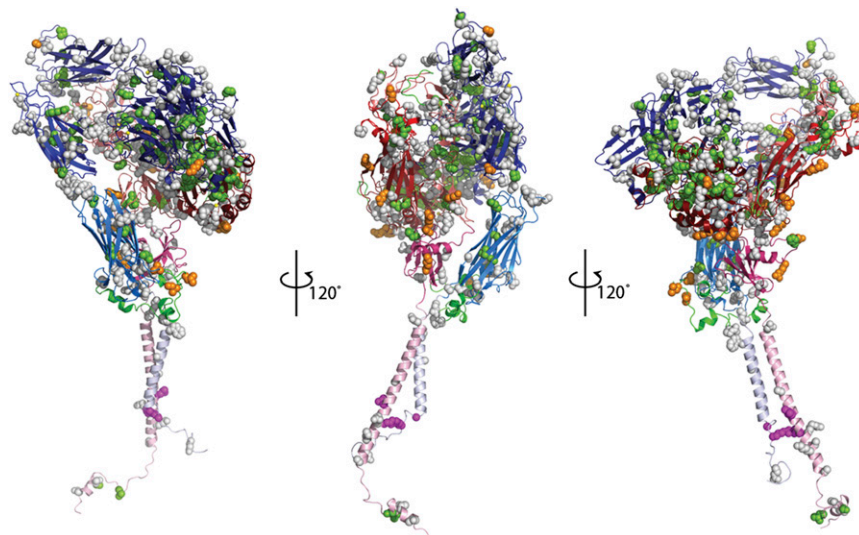
**Glanzmann Thrombasthenia.** The ThromboGenomics  $\alpha$ Ib $\beta$ 3 compendium developed from the five databases (HGMD, 1000G, U.K.10KWES, U.K.10KWGS, and ESP) contains 319 missense variants in either  $\alpha$ Ib or  $\beta$ 3, and these are reported in [Table S1 A and B](#). Putative mutations producing GT were considered to be confirmed as disease-causing if they were associated with a GT phenotype and the negative impact of the mutation on  $\alpha$ Ib $\beta$ 3

expression, activation, and/or ligand binding was established in a recombinant system based on a review of the publications describing each of the mutations. Mutations reported in association with a GT phenotype that were not studied in a recombinant system were considered presumptive GT mutations. Because some patients with GT have at least one allele with no exonic variant, it is possible that a nondeleterious (silent) variant that just happened to be on such an allele will be mistaken for a GT-causing variant. This possibility is highlighted by the identification of a silent variant with a relatively high frequency (1%) on an allele in a patient with GT that had no other exonic variants (60). This variant might have been misidentified as disease-causing if its functional consequences were not analyzed in a recombinant expression system.

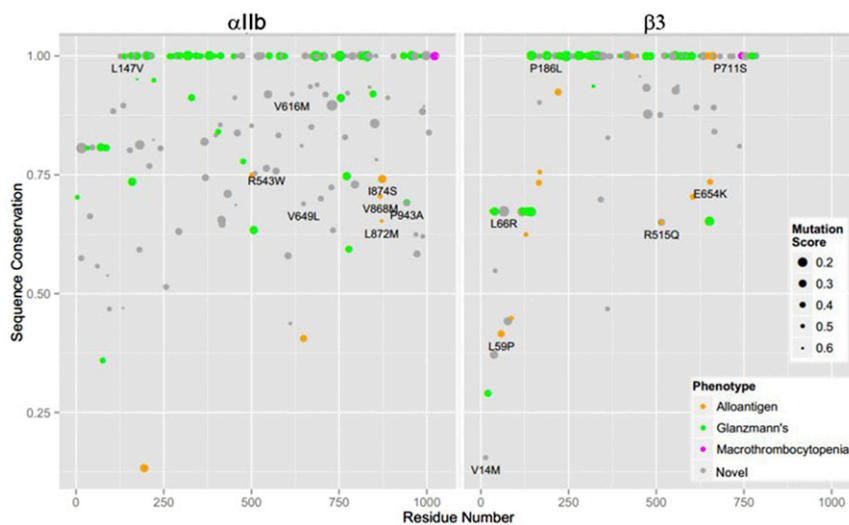
A total of 111 GT missense mutations were identified (56 in  $\alpha$ Ib and 55 in  $\beta$ 3), of which 52 were considered confirmed and 59 presumptive. The locations of these mutants on the complete  $\alpha$ Ib $\beta$ 3 molecular model are indicated by green spheres in [Fig. 1](#) and [Movie S1](#). For each of these mutations, we analyzed the degree of conservation of the substituted amino acid across species (based on a scale of 0–1, with 1 being the most conserved). A mutation score was then assigned to each missense mutation as an indication of the severity of the amino acid change (based on a scale from 0 to 1, with 0 being the most severe), that is, taking into account the likely amino acid substitutions during evolution. The results are shown in [Fig. 2](#), with green dots indicating the GT mutants, and the different dot sizes indicating the severity of the variant. The majority of disease-causing GT missense mutations involve highly conserved amino acids (scores at or near 1.0) and severe amino acid changes that are less tolerated evolutionarily (55% of variants responsible for GT have mutation scores <0.4).

It is notable that not one of the established GT mutations was identified on any of the up to ~32,000 alleles analyzed in the four WES and WGS studies represented in the databases, suggesting that all of these GT mutations have allele frequencies less than 0.01% in the studied populations.

**Alloantigens.** The recognized alloantigens on  $\alpha$ Ib and  $\beta$ 3 are listed in [Table S2](#) and shown as orange spheres in [Fig. 1](#) and on the complete  $\alpha$ Ib $\beta$ 3 molecular model in [Movie S1](#). The  $\alpha$ Ib HPA-3b/Bak<sup>a/b</sup> (I874S) alloantigen is by far the most common polymorphism, being found in 37% of the alleles and identified in individuals in all five databases. Three of the other reported



**Fig. 1.** Model of  $\alpha$ Ib $\beta$ 3 with Glanzmann thrombasthenia mutations (green spheres), macro/anisothrombocytopenia mutations (magenta spheres), alloantigens (orange spheres), and novel missense variants (gray spheres).



**Fig. 2.** Analysis of the degree of conservation of the substituted amino acids across species (based on a scale of 0–1, with 1 being the most conserved). A mutation score is also assigned to each missense variant as an indication of the severity of the amino acid change (based on a scale from 0 to 1, with 0 and the larger dot size indicating the most severe variant), that is, taking into account likely amino acid substitutions during evolution. Circles identified as 0.2 represent mutations scores between 0–0.2, those identified as 0.3 represent mutation scores between 0.2–0.3, and so forth.

$\alpha$ IIb alloantigens were identified in at least one of the four WES and WGS studies in addition to HGMD, and they had minor allele frequencies (MAFs) of 0.4% (HPA-27b/Cab3, Ak-L872M), 0.2% (HPA-9b/Max-V868M), and 0.03% (HPA-20b/Kno-T650M). Two of the reported  $\alpha$ IIb alloantigens (HPA-24b/Cab2, In-S503N and HPA-22b/She, Sey-K195T) were not identified in any of the four studies and thus presumably have MAFs  $< 0.01\%$  in the studied populations. HPA-1b/PI<sup>A2</sup>-L59P was the most common  $\beta$ 3 alloantigen with an MAF of 14%. The other  $\beta$ 3 alloantigens had MAFs of 0.9% or less and only HPA-6b/Ca/Tu<sup>a</sup>-R515Q was identified in all four WES and WGS studies. When analyzed on the basis of amino acid conservation, compared with GT missense mutations, the variants responsible for the alloantigens (orange dots in Fig. 2) were less likely to affect highly conserved amino acids or to exhibit severe amino acid changes (40% of variants responsible for alloantigens have a mutation scores  $< 0.4$ ; differences in conservation score values between GT and alloantigens:  $P = 9 \times 10^{-5}$ ; difference in location = 0.25; Wilcoxon rank sum test).

**Macro/Anisothrombocytopenia.** Five heterozygous  $\alpha$ IIb $\beta$ 3 missense mutations affecting four different amino acids ( $\alpha$ IIb G1022C, R1026Q, and R1026W;  $\beta$ 3 L744P and D749H), along with three different deletions, have been associated with macrothrombocytopenia or platelet anisocytosis and variable abnormalities in platelet function and clinical hemorrhage (32–38). The corresponding positions on the complete  $\alpha$ IIb $\beta$ 3 molecular model are shown as magenta spheres in Fig. 1 and [Movie S1](#). All of the mutations appear to result in constitutive activation of the receptor, which in turn impairs proplatelet formation. None of these mutations was identified in the four WES and WGS databases, suggesting that they, like the mutations causing GT, are likely rare (MAF  $< 0.01\%$ ). The conservation and mutation scores for these amino acids are shown in Fig. 2.

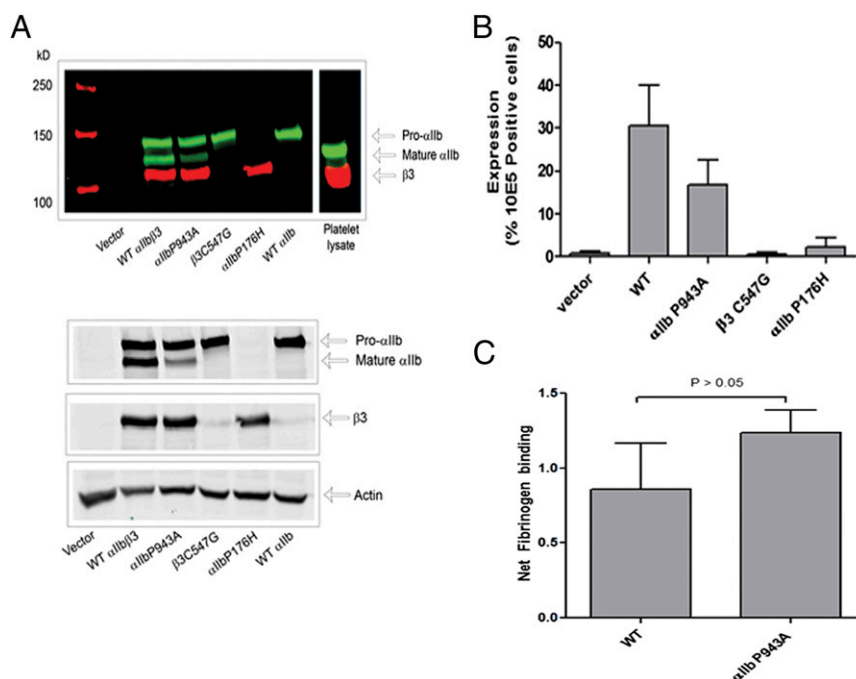
**Novel Variants.** A total of 114 novel missense variants in  $\alpha$ IIb and 68 novel missense variants in  $\beta$ 3 were identified. These are shown as gray spheres in Fig. 1 and on the complete  $\alpha$ IIb $\beta$ 3 molecular model in [Movie S1](#). In total, 11.2% of the 1,025 amino acids in mature  $\alpha$ IIb and 8.7% of the 784 amino acids in mature  $\beta$ 3 were mutated. Only 3.8% of the variants had MAFs  $> 0.1\%$ , indicating their relative rarity, consistent with the expectation under the Fisher-Wright model (61) that most variants are rare. Six  $\alpha$ IIb amino acids were

mutated twice (40 T/I and T/A, 369 H/Q and H/D, 433 R/C and R/H, 543 R/W and R/Q, 728 A/T and A/V, and 988 R/W and R/Q) as were two  $\beta$ 3 amino acids (14 V/L and V/M and 186 P/S and P/L) among the novel variants and the previously described incidental nondisease-causing variants. The severity of the amino acid change and the conservation across species of the reference amino acid of each variant are shown in Fig. 2 (gray dots). Compared with the GT missense mutations, there was greater representation of variants at less conserved amino acids and, in general, the variants produced less severe amino acid changes (30% of the novel variants have mutation scores  $\leq 0.4$ ).

**Assessment of the Effect of Three Novel Variants on  $\alpha$ IIb $\beta$ 3 Expression and Function.** Five of the novel variants in our study affected an amino acid in  $\alpha$ IIb previously reported to be associated with GT (P176H, L214F, H813L, L830F, and P943A) and two affected an amino acid in  $\beta$ 3 previously reported to be associated with GT (R131W and C547G). The allele frequency for P943A was 0.46%, whereas it was 0.01% or less for the other variants. It is notable that  $\alpha$ IIb P176 appears to be a mutational hotspot because four patients from three ethnically distinct kindreds (Mennonite, Dutch, and Italian) have been reported to have P176A mutations and a Chinese GT patient has been reported with a P176L mutation (62, 63).

We assessed the effect of the three variants with the most severe variant scores,  $\alpha$ IIb P176H and P943A, and  $\beta$ 3 C547G, by expressing them in HEK293 cells and analyzing both receptor expression and the ability of the receptor to bind fibrinogen when activated with the reducing agent DTT (53). Fig. 3A shows the immunoblot data from the cell lysates. The WT  $\alpha$ IIb $\beta$ 3 bands correspond to pro $\alpha$ IIb, mature  $\alpha$ IIb (which aligns with platelet  $\alpha$ IIb), and  $\beta$ 3 (which migrates slightly slower than platelet  $\beta$ 3 because it contains a V5 tag). The  $\alpha$ IIb P943A variant expressed similar amounts of pro $\alpha$ IIb, but reduced amounts of mature  $\alpha$ IIb, suggesting a partial defect in  $\alpha$ IIb $\beta$ 3 maturation. In contrast, the  $\alpha$ IIb P176H variant showed a dramatic reduction in pro $\alpha$ IIb expression and virtually no mature  $\alpha$ IIb, and the  $\beta$ 3 C547G variant resulted in a profound decrease in  $\beta$ 3 expression. Flow cytometry using an antibody specific for the  $\alpha$ IIb $\beta$ 3 complex (mAb 10E5) confirmed that the P943A variant produced a partial reduction in  $\alpha$ IIb $\beta$ 3 expression and that the other two variants produced profound reductions in surface expression of the  $\alpha$ IIb $\beta$ 3 complex





**Fig. 3.** (A) HEK293 cells ( $10^5$  cells/mL) were transfected with cDNA harboring the indicated variant in  $\alpha$ IIb or  $\beta$ 3. Cell lysates were subject to immunoblotting with antibodies to  $\alpha$ IIb and  $\beta$ 3. Immunoblot shown is representative of five experiments. (Upper) Color image showing detection with Odyssey system. (Lower) Black and white image of the examined proteins. (B) Transfected HEK293 cells were analyzed by flow cytometry for  $\alpha$ IIb $\beta$ 3 expression. Cells were incubated with Alexa488-conjugated mAb 10E5 (10  $\mu$ g/mL) or (as a control for nonspecific binding) 10  $\mu$ g/mL Alexa488-conjugated mAb 10E5 in the presence of excess (125  $\mu$ g/mL) unlabeled 10E5. The graph depicts the mean  $\pm$  SEM;  $n = 5$ . (C) To assess  $\alpha$ IIb $\beta$ 3 function, DTT (10 mM) was added to the cells and Alexa647-conjugated fibrinogen binding was assessed via flow cytometry. Binding is expressed as net fibrinogen binding. Data are expressed as mean  $\pm$  SEM;  $n = 5$ .

(Fig. 3B). The cells expressing the  $\alpha$ IIb P943A variant were then further tested for their ability to bind fibrinogen in the presence of DTT. When the amount of fibrinogen bound was normalized for  $\alpha$ IIb $\beta$ 3 expression, the  $\alpha$ IIb P943A variant was found to bind as much fibrinogen as normal  $\alpha$ IIb $\beta$ 3 (Fig. 3C), suggesting that this variant partially affects expression, but does not alter receptor function.

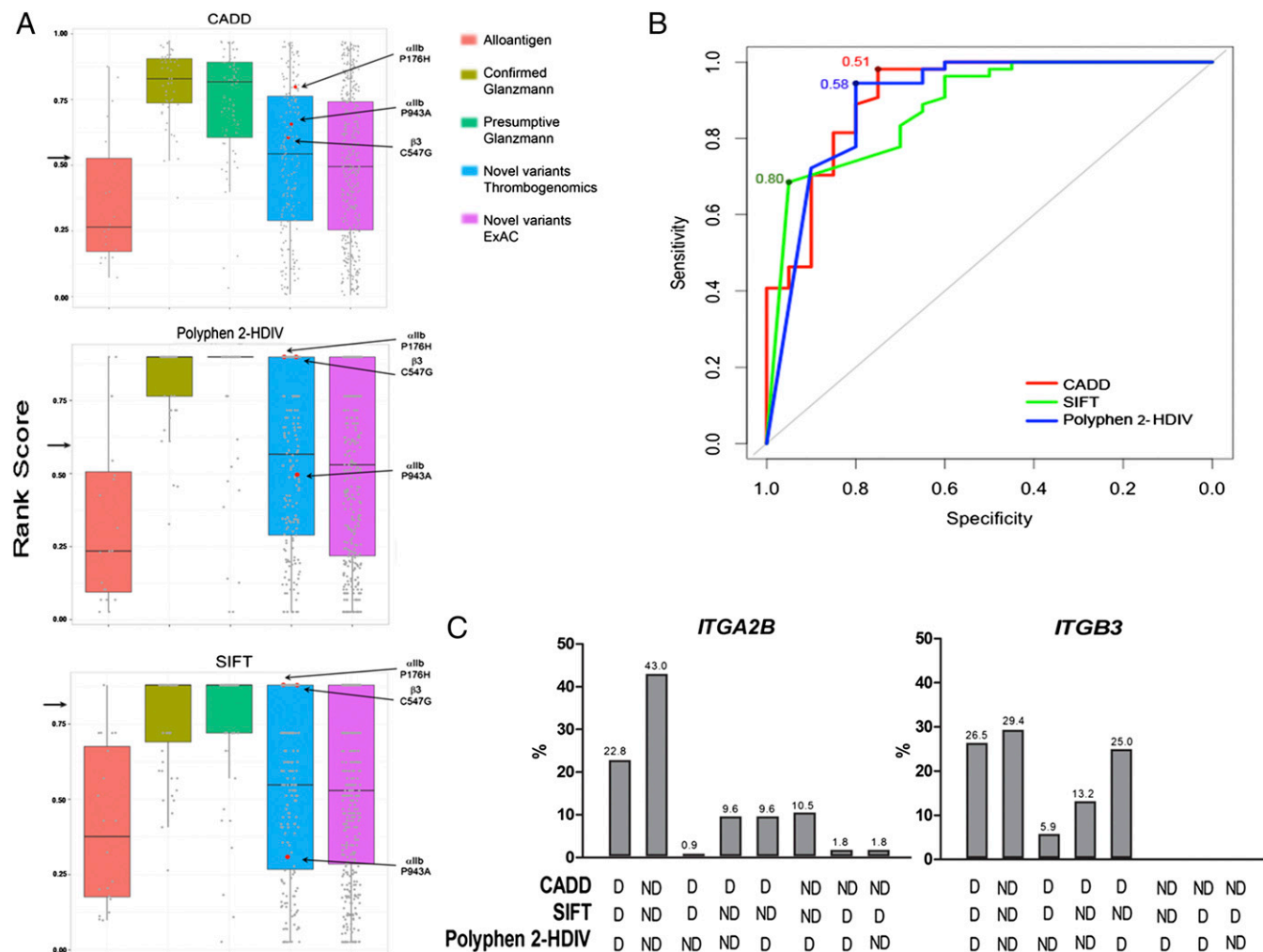
#### Novel Variants Affecting Amino Acids Mutated in Reported Alloantigens.

We identified two novel variants affecting the same amino acid responsible for an alloantigen,  $\beta$ 3 R515W affecting the same amino acid as HPA-6b (R515Q) and  $\beta$ 3 R662H affecting the same amino acid as HPA-11b (R662C). Both novel variants were identified in only one of the studies. R515W has an allele frequency of 0.1% among the 32,216 alleles analyzed, whereas the alloantigen R515Q has an allele frequency of 0.01% based on analysis of 13,306 alleles. R662H has the same allele frequency as the alloantigen variant R662C (0.02%).

**Predicting the Consequences of Individual Variants.** A number of algorithms have been developed to predict whether a given amino acid substitution is likely to be deleterious—that is, whether it will alter expression or function. We used three different commonly used algorithms: combined annotation-dependent depletion (CADD) (57), Polyphen 2-HDIV (58), and sorting intolerant from tolerant (SIFT) (59), and analyzed the missense variants based on whether they were confirmed GT mutants, presumptive GT mutants, alloantigens, or novel variants. We anticipated that the confirmed and presumptive GT mutants would be judged to have the highest likelihood of being deleterious, the alloantigens would have the lowest likelihood, and the novel variants would have the greatest range of likelihoods. The results varied with the prediction tool, but in general were consistent with our expectations (Fig. 4A). Moreover, each of the algorithms identified a cluster of

three presumptive GT variants with a low probability of being deleterious, and it may be valuable to study these in more detail [ $\alpha$ IIb P76A and  $\beta$ 3 M321L (64), and H652L (65)]. The concordance between prediction tools for individual variants is displayed as a heat map in Fig. S1 and demonstrates variable degrees of discordance among the tools.

To obtain an estimate of the sensitivity and specificity of each of the prediction tools in judging missense variants, we assessed their sensitivity in identifying the confirmed GT and macro/anisothrombocytopenia mutations as deleterious (positive controls) and their specificity in identifying the 20 alloantigens as not being deleterious (negative controls). We then used the results to create receiver operating characteristic (ROC) curves for each of the prediction tools (Fig. 4B). All three gave similar patterns, with areas under the curves of 0.91 [median; 95% confidence intervals (CI) 0.83, 0.99], 0.90 (CI 0.81, 0.99), and 0.88 (CI 0.80, 0.97) for CADD, Polyphen 2-HDIV, and SIFT, respectively (Table 1) ( $P =$  not significant for all pair-wise comparisons). To estimate the percentage of the novel variants we identified that are likely to be deleterious, we applied the cut-off value of each prediction tool that best differentiated the positive controls from the negative controls according to the ROC curve and then applied this value to the novel variants. These optimal cut-off values, achieved median sensitivities of 98%, 94%, and 69% and specificities of 75%, 80%, and 95% for CADD, Polyphen 2-HDIV, and SIFT, respectively. The percentages of novel  $\alpha$ IIb missense variants predicted to be deleterious were 43.0%, 45.6%, and 27.2% for CADD, Polyphen 2-HDIV, and SIFT, respectively, and the comparable values for  $\beta$ 3 missense variants were 70.6%, 54.4%, and 32.4%, respectively. Because it is possible that some of the low-frequency alloantigens are deleterious, we also created ROC curves using only the 11 of 20 alloantigens that were judged by all three algorithms to have a low probability of being deleterious (rank scores  $< 0.5$ ) as negative controls (Table S1A). As expected, the areas under the



**Fig. 4.** (A) A prediction algorithms to assess the likelihood that a variant will be deleterious. The results of the analysis with the algorithms CADD, SIFT, and Polyphen 2-HDIV are presented for the alloantigens, the presumptive and confirmed GT variants, and the ThromboGenomics and Exome Aggregation Consortium (ExAC) novel variants. Black arrows on ordinate denote optimal rank score cut-off values. (B) ROC curves of data based on considering confirmed GT and macro/anisothrombocytopenia as true positives and 20 alloantigens as true negatives. The optimal discriminator points are identified by dots and the corresponding rank scores are shown next to the dots. (C) Concordance of algorithms in predicting whether novel variants are likely to be deleterious using the optimal cut-offs determined by the ROC curves. D, deleterious; ND, nondeleterious.

curves were higher (1.00, 0.99, and 0.99, respectively). The optimal cut-off values for each algorithm yielded sensitivities of 98%, 94%, and 96%, respectively and 100% specificity for all three (Table 1). The percentages of novel  $\alpha$ Ib missense variants predicted to be deleterious based on these optimal cut-off points were 49.1%, 47.4%, and 59.7%, respectively; the comparable values for  $\beta$ 3 were 73.5%, 59.9%, 67.7%, respectively. Thus, only SIFT showed a major change in the prediction of deleteriousness, and this reflects the more step-wise nature of the prediction rank scores with this algorithm in comparison with the more continuous rank scores with the other algorithms (Fig. 4A).

**Concordance in Predictions Among the Three Algorithms.** The concordance among the three different algorithms in classifying the novel variants is shown in Fig. 4C. All three algorithms classified 22.8% of the *ITGA2B* and 26.5% of the *ITGB3* variants as deleterious and 43.0% of the *ITGA2B* and 29.4% of the *ITGB3* variants as nondeleterious. Thus, in total, there was concordance among all three in classifying 65.8% of the *ITGA2B* variants and 55.9% of the *ITGB3* variants. The greatest discordance was found in the classification of *ITGB3* variants as being deleterious by

both CADD and Polyphen 2-HDIV, but nondeleterious by SIFT, which accounted for 25.0% of the novel *ITGB3* variants.

It is of particular note that Polyphen 2-HDIV and SIFT identified the  $\alpha$ Ib P176H and  $\beta$ 3 C547G variants that we found to have profound defects in surface expression as nearly certain to be deleterious, with rank scores much higher than the optimal cut-off values. In contrast, CADD gave both lower rankings (0.66 and 0.61, respectively), but still ones above the optimal cut-off value of 0.51. As a result, all three algorithms classified these variants as deleterious. In contrast, the rank scores given to the  $\alpha$ Ib P943A variant were more broadly distributed, with SIFT giving it the lowest ranking (0.31), Polyphen 2-HDIV an intermediate ranking (0.50), and CADD the highest ranking (0.80). Based on each of their optimal cut-off values, this would result in SIFT and Polyphen 2-HDIV classifying this mutation as nondeleterious and CADD ranking it as deleterious. These results are particularly interesting because we found that the variant partially decreased  $\alpha$ Ib $\beta$ 3 expression.

**Analysis of the Exome Aggregation Consortium Database.** After our studies were completed, the Exome Aggregation Consortium made available the results of 60,706 unrelated individuals whose DNA

**Table 1. ROC analysis of  $\alpha$ IIB $\beta$  variants**

Algorithm	AUC	AUC Confidence Interval	Sensitivity	Specificity	Cut-off	Percentage of novel variants predicted to be deleterious	
						<i>ITGA2B</i>	<i>ITGB3</i>
Specificity based on all 20 alloantigens							
CADD	0.91	0.83–0.99	0.98	0.75	0.51	43.0	70.6
SIFT	0.88	0.80–0.88	0.69	0.95	0.80	27.2	32.4
Polyphen-2	0.90	0.81–0.90	0.94	0.80	0.58	45.6	54.4
Specificity based on 11 alloantigens with low probability of being deleterious							
CADD	1.00	0.99–1.00	0.98	1	0.46	49.1	73.5
SIFT	0.99	0.98–0.99	0.96	1	0.44	59.7	67.7
Polyphen-2	0.99	0.98–0.99	0.94	1	0.55	47.4	55.9

AUC, area under the ROC curve.

was sequenced as part of various disease-specific and population genetic studies. We analyzed the 319 novel variants (that is, those remaining after removing the GT and macro/anisothrombocytopenia mutations and the alloantigens) in  $\alpha$ IIB $\beta$  (72 of which are also reported in the ThromboGenomics dataset) and the 216 novel variants in  $\beta$ 3 (49 of which are also reported in the ThromboGenomics dataset) using the optimal cut-off values for each algorithm based on the ROC curves. CADD, Polyphen 2-HDIV, and SIFT predicted that 40.8%, 35.7%, and 23.5%, respectively, of the  $\alpha$ IIB and 60.6%, 46.3%, and 29.6%, respectively, of the  $\beta$ 3 variants would be deleterious (Fig. 4A), and all three predicted that 17.6% of  $\alpha$ IIB and 22.7% of the  $\beta$ 3 variants would be deleterious, which are values similar to those predicted using the ThromboGenomics dataset.

## Discussion

Our study highlights the opportunity and challenges in interpreting the potential impact of novel variants of  $\alpha$ IIB $\beta$  on receptor function, with broader implications for interpreting novel variants in other genes and gaining insights into the evolutionary pressures that have affected  $\alpha$ IIB $\beta$ . We found that three different prediction algorithms were able to differentiate confirmed GT and macro/anisothrombocytopenia mutations from alloantigens with reasonably high sensitivity and specificity, but there was some overlap between the groups and the three algorithms were only about 60% concordant in classifying novel variants. Thus, it is important to augment the results of these analyses with additional data. Table S3 summarizes some of the considerations that a clinician can bring to bear on deciding whether a novel variant in  $\alpha$ IIB or  $\beta$ 3 is likely to result in a medically meaningful phenotype. If a propositus is available for study,  $\alpha$ IIB $\beta$ 3 can be quantified by flow cytometry and its function assessed directly by platelet aggregation studies using ADP or other agonists that rely on the  $\alpha$ IIB $\beta$ 3 receptor. Individuals who are heterozygous for mutations that affect expression of the receptor usually have about 50–60% of the normal level of platelet  $\alpha$ IIB $\beta$ 3 expression, but there is some overlap between the values of heterozygotes and normals (66). Homozygotes have markedly decreased platelet aggregation, but heterozygotes have little or no decrease in platelet aggregation. If the propositus is not available, as might occur if a fetus is tested, data from the prediction tools can be combined with additional data, including whether the variant affects an amino acid previously associated with causing GT or macro/anisothrombocytopenia, or whether the amino acid is highly conserved or is in a region known to have very stringent structural requirements (for example, the blades of the  $\beta$ -propeller in  $\alpha$ IIB) or to affect ligand binding.

If molecular modeling or molecular dynamics simulations are available, they can provide more detailed information on the likely impact of the variant. Additional information can come from producing a recombinant version of the variant in a cell line and testing expression and ligand binding, but this requires considerable time and research expertise, and there are regulatory issues in providing information for diagnostic purposes from assays that are not performed in certified laboratories. The MAF, if known, can also provide valuable information, because one would expect a deleterious variant with a high MAF to have already come to medical attention. By combining all of this information, the clinician will be able to make a reasoned judgment, but in many cases there will still be an irreducible level of uncertainty that needs to be carefully conveyed to the individual requesting the information.

The results from recent analyses of human exomes and genomes in large populations provides context for interpreting our data. The explosive growth in the human population over the past 2,500 y, estimated to be from ~50 million to ~7 billion people, implies departure from population genetic equilibrium and predicts a large numbers of relatively new variants that have had little time to undergo selective evolutionary pressure (2, 3, 67, 68). Because they entered the gene pool recently, these variants are expected to have a low allele frequency, and vice versa, unless the variant confers a strong negative effect on fertility in the heterozygous state, one can surmise that variants with a low MAF entered the gene pool recently and thus are relatively new. Variants with an MAF < 0.01% probably originated within the past 250 y and those with an MAF of ~0.5% probably originated ~2,500 y ago (68). By definition, one must sequence large numbers of individuals with high precision to robustly establish a low MAF, and so such data have only become available in the past few years (3). Racial differences also need to be considered, because the African population is the oldest and thus most diverse, and the European population, which is presumed to have undergone two population bottlenecks, probably has had the most rapid recent population growth (2, 3). Establishing that rare variants are not the result of sequencing errors using current methodology requires careful attention, and new methods to minimize the risk of mistaking sequencing errors for variants are being developed (69).

It is important to recognize the limitations of our study. In particular, the number of variants identified may be underestimated if the sequencing does not include all of the exons, as was the case for several exons in the  $\beta$ 3 subunit gene in the ESP study, representing ~40% of the exonic bases (Fig. S2), or if the filters are stringent in the face of relatively low coverage. On the other



hand, if there is low sequencing coverage of a region, as was the case for one complete exon and a portion of another exon in the  $\alpha$ Ib subunit gene in the ESP study, and in the low-coverage U.K.10KWGS study (70), it is possible that a sequencing error may be misreported as a variant. To minimize this risk, a series of filters are applied to the raw sequencing data before making a final determination of whether a variant is present. In the case of the U.K.10KWGS study, we found it reassuring that the percentage of individuals with missense mutations (1.27%) was similar to the combined value of 1.39% for the other three studies. Moreover, 37 of the 48 missense variants identified in the U.K.10KWGS were also identified by at least one other study. Overestimation of variant frequency may also occur if one uses cultured immortalized lymphoblastoid cells as the source of DNA because they acquire variants during culture that are not present in the germ line genome (71). Lymphoblastoid cells were used as a source of DNA only in the 1000G database. Thus, with the above caveats, we believe our dataset provides a reasonable reflection of the missense variants in the populations studied. We did not include variants involving DNA base pair deletions or insertions because these usually are much less common than missense variants, and next-generation sequencing is less accurate in detecting them.

Because not one of the reported GT missense mutants reported in the literature (HGMD) (Table S1B) was identified in the alleles included in this analysis, they all presumably have very low MAFs, implying that they entered the population recently. Independent support for the recent origin of GT missense mutations comes from studies of two populations with relatively high rates of GT because of intragroup marriage, namely Israeli and Palestinian Arabs and French Manouche gypsies, in which the founder mutations were estimated to be only ~300–600 and ~300–400 y old, respectively (23, 24).

Against this background, our data provide new insights into the frequency and potential impact of novel variants of  $\alpha$ Ib $\beta$ 3. In the ~16,500 individuals studied, we found 114 novel missense variants affecting ~10% of the amino acids in  $\alpha$ Ib and 68 novel missense variants affecting ~9% of the amino acids in  $\beta$ 3. Thus, with the limitations noted above, we found that ~1.3% of the population studied harbors a variant in either  $\alpha$ Ib or  $\beta$ 3.

$\alpha$ Ib $\beta$ 3 is expressed at ~100,000 copies per platelet, which is approximately twice the density required for normal platelet aggregation (72, 73). As a result, individuals who are heterozygous for GT variants are usually asymptomatic or have relatively mild symptoms (66, 74–76), and thus are unlikely to be subjected to selective pressure. Because hemorrhage can result in death before reaching sexual maturity or during or after child-birth, it is possible that there would be selective pressure against  $\alpha$ Ib or  $\beta$ 3 variants that decreased receptor expression or function in mammals if they had a lower density of  $\alpha$ Ib $\beta$ 3 receptors than modern humans or had a high percentage of intragroup matings. The lack of GT mutations with high MAFs is consistent with selective pressure against deleterious variants in the distant past. Platelets are known to be present in all mammals, including monotremes (which lay eggs) and marsupials (which do not have placental births), and so the hemorrhagic risk associated with placental births cannot explain the appearance of platelets in mammals (77). To our knowledge however, the densities of  $\alpha$ Ib $\beta$ 3 receptors in monotreme and marsupial platelets have not been reported, and so it remains possible that placental birth selected for high receptor density. For comparison, the density of  $\alpha$ Ib $\beta$ 3 receptors on chicken thrombocytes has been reported to be only ~5% of that of human platelets (78). Because humans increasingly mate with individuals outside their kinship, heterozygous deleterious variants are likely to continue to accumulate. Such variants may become more consequential in the future as they spread through the population. Additional variants are likely to enter the population as de novo variants. Individuals are estimated to have an average of ~74 de novo single-nucleotide variants (79);

with a world population of ~7 billion people, one can expect more than 500 billion variants per generation in the “humanome,” enough on average, to affect each of the 3 billion genome bases more than 100-times over. Thus, with time these variants will increase the likelihood of GT because of double heterozygosity.

Our studies of  $\alpha$ Ib $\beta$ 3 serve as a paradigm for the opportunities and challenges one can anticipate as we enter the era in which many individuals (and fetuses) undergo whole-exome and whole-genome sequencing and want to know whether a novel variant is deleterious. We applied three algorithms that are currently used to estimate the percentage of novel variants that are deleterious and found that using their optimal ROC rank cut-off values, they predicted between 27.2% and 45.6% of the novel  $\alpha$ Ib variants and between 32.4% and 70.6% of the novel  $\beta$ 3 variants are deleterious; all three predicted that 22.8% of  $\alpha$ Ib variants and 26.5% of  $\beta$ 3 variants would be deleterious. These percentages can be compared with the data from the ESP study (2), where 74% of nonsynonymous variants (98% of which were missense variants) were judged deleterious by at least one of seven algorithms, but only 27.9% met the criterion of being predicted deleterious by at least four of the seven algorithms. We would emphasize, however, that the optimal ROC cut-off value is based on discriminating between two groups, with sensitivity and specificity given equal weight. Clinically, however, one may wish to set a minimum sensitivity or specificity to address a specific goal, for example, choosing a high-sensitivity value so as to avoid a false-negative diagnosis. This can be achieved by identifying the rank cut-off value corresponding to the desired sensitivity/specificity point on the ROC curve.

Two of the novel variants we identified ( $\alpha$ Ib P176H and  $\beta$ 3 C547G) were predicted by all three algorithms to be deleterious and we found that both profoundly affected  $\alpha$ Ib $\beta$ 3 expression. We anticipated that the  $\alpha$ Ib P176H variant would decrease expression because of its strategic location and the major change in the side-chain structure, and that the  $\beta$ 3 C547G variant would decrease receptor expression because of its impact on disulfide bonding. The  $\alpha$ Ib P943A variant was predicted by two of the three algorithms to be deleterious and we found that it partially decreased  $\alpha$ Ib $\beta$ 3 expression and did not impair fibrinogen binding induced by DTT. We suspected that this latter variant would not cause a severe defect in  $\alpha$ Ib $\beta$ 3 expression or function because, with its relatively high MAF (0.46%), one would have expected it to have already come to medical attention in homozygous or double-heterozygous GT patients. It is possible that hypomorphic variants like P943A, which partially reduce  $\alpha$ Ib $\beta$ 3 expression, account for some of the wide range of  $\alpha$ Ib $\beta$ 3 expression observed among apparently healthy individuals (66). Thus, variants like P943A highlight that even with the availability of multiple prediction algorithms and an advanced understanding of a protein's structure and function, it may be challenging to predict the functional consequences of any given variant.

Similar considerations bear on the ability to predict whether a given variant will result in the generation of an alloantigen. Despite advances in molecular modeling of platelet alloantigens (47) and predicting the immunogenicity of HLA variants (80) ([www.hlamatchmaker.net/](http://www.hlamatchmaker.net/)), and the ability of peptides to bind to specific major histocompatibility molecules (81), it is challenging to predict the likelihood that a mismatch between parents in their sequences of  $\alpha$ Ib or  $\beta$ 3 will result in medically meaningful fetal and neonatal alloimmune thrombocytopenia because many other immunologic factors may influence the outcome (82). Whether  $\beta$ 3 variants induce antibodies that affect  $\alpha$ V $\beta$ 3 function, potentially leading to intrauterine growth restriction or vascular abnormalities contributing to hemorrhage, or whether de novo  $\alpha$ Ib or  $\beta$ 3 variants can result in constitutively active receptors that may lead to thrombus formation in the placenta (82), are additional questions of import. It is possible that the latter abnormalities have not come to medical attention to date because they result in fetal loss.

In conclusion, next-generation sequencing, coupled with advanced bioinformatics algorithms, is making it possible to both identify novel missense variants and in many cases make a reasoned estimate of the likelihood that the variant will affect protein expression and/or function. Current algorithms are not definitive, however, in a substantial percentage of cases, and so these tools need to be used judiciously and coupled with additional clinical, laboratory, and population-based data.

## Methods

**Variant Analysis.** This study is part of the ThromboGenomics working group (see *SI Methods* for description). Data for the present study on  $\alpha$ IIb and  $\beta$ 3 variants were obtained from the whole-exome and whole-genome sequencing data from five different databases: (i) HGMD ([www.hgmd.cf.ac.uk/ac/index.php](http://www.hgmd.cf.ac.uk/ac/index.php)), which collates published gene lesions responsible for human inherited disease; (ii) The 1000 Genomes project (1,000GP; [www.1000genomes.org](http://www.1000genomes.org)), representing 1,092 individuals of East Asian, South Asian, African, European, and American ancestry who underwent whole exome or low pass whole genome sequencing (83); (iii) The U.K.10KWES ([www.uk10k.org](http://www.uk10k.org)), representing up to 4,732 members of the British population (77); (iv) The U.K.10KWGS, representing 3,781 members of the British population; and (v) The National Heart, Lung and Blood Institute ESP, representing up to 6,503 Americans of European and African ancestry (84). We restricted our analyses to single-nucleotide variants that lead to amino acid substitutions (missense variants). The variants were expressed in a variant call format, starting with the first based in the human genome reference sequence, which includes the signal peptide. As a result, the numbers are 31 amino acids higher than those for mature  $\alpha$ IIb and 26 higher than those for mature  $\beta$ 3. The effects of variants were predicted on the standard *ITGA2B* reference transcript ENST00000262407 (NM\_000419.3, LRG\_479t1) and the *ITGB3* transcript ENST0000055948 (NM\_000212.2, LRG\_481t1) using the variant effect predictor.

Data on coverage of the exons in *ITGA2B* and *ITGB3* are available from the ESP study and the results shown in Fig. S2. All of the exons in *ITGA2B* had more than 15 $\times$  coverage, except for the 5' end of exon 2 and exon 27. Exons 1–10 in *ITGB3* had greater than 15 $\times$  coverage, but exons 11–15 were not analyzed. In the U.K.10KWGS study, sequencing was performed to a median depth of 7.6 reads per sample at single-nucleotide variant sites (70). Each study defined a series of criteria by which they validated the variants they identified (83, 84) ([www.uk10k.org/data.html#datasets](http://www.uk10k.org/data.html#datasets)).

**Catalogue of GT Missense Mutants.** We combined data from the online registry of GT variants and the five databases identified above into Table S1, where part A of the table includes data from HGMD only and part B of the table includes combined data from 1000G, U.K.10KWES, U.K.10KWGS, and ESP. All reported variants where reviewed and those in which the proposed GT mutant was shown to lead to a defect in receptor expression and function in a recombinant system were categorized as confirmed GT mutations. All other variants where considered presumptive GT mutations. Defects in GT patients were also characterized as either homozygous or heterozygous if the information was available. The MAF of each missense variant is reported based on the combined data from the different databases. Finally, the database in which the variant was reported is identified.

**Analysis of Missense Variants.** A complete  $\alpha$ IIb $\beta$ 3 3D model was built using a combination of X-ray crystallographic data of the  $\alpha$ IIb $\beta$ 3 ectodomain [PDB ID code 3FCS (18)] and NMR data of the transmembrane and cytoplasmic domains [PDB ID codes 2K9J (85), 2KNC (86), and 2KV9 (87)]. Missing seg-

ments of the  $\alpha$ IIb calf-2 domain (amino acid residues 763–775 and 839–874), the  $\beta$ 3 hybrid domain (amino acid residues 74–79), and the link between the EGF-1 and EGF-2 domains (amino acid residues 476–483) were built ab initio using Rosetta 2.2 (88) (Movie S1).

Missense variants were analyzed in terms of: (i) the level of conservation of the residue position in the alignment of the wild-type sequence of  $\alpha$ IIb or  $\beta$ 3 with sequences from different species (the ordinate of the graph in Fig. 2), and (ii) the severity of the specific amino acid change (the size of each circle in Fig. 2).

Both the level of conservation across different species and the severity of the missense variant were assessed based on a mutation score calculated as the “sum-of-pairs” score with a modified PET91 substitution matrix (89) by Valdar and Thornton (see ref. 54 and equation 37 in ref. 55) using the Scorecons server at [www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons\\_server.pl](http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons_server.pl). Severe mutations are less tolerated during evolution and are therefore characterized by low scores in the substitution matrix, as opposed to mutations to amino acids with similar properties. Large circle sizes in Fig. 1 (variant score = 1 – “sum-of-pairs” score) were chosen to represent severe variants.

In the calculation of sequence conservation, the mutation score consists of the average of the pair scores calculated for each amino acid in an aligned column, weighted using the dissimilarity between the sequences comprising each pair to normalize against sequence redundancy in the alignment. Sequence alignments of all human and vertebrate  $\alpha$ IIb or  $\beta$ 3 sequences were obtained using the multiple sequence alignment Hidden Markov Model/clustal W method (56).

**Mutagenesis, Cell Transfection, and  $\alpha$ IIb $\beta$ 3 Analysis.** For mutagenesis, cell transfection, and  $\alpha$ IIb $\beta$ 3 analysis, see *SI Methods*.

**Prediction of Variant Impact.** We curated the variants using the dbNSFP v2.4 database (database for nonsynonymous SNP's functional predictions; <https://sites.google.com/site/jpopgen/dbNSFP>). This database integrates prediction scores from several different algorithms, including CADD (57), Polyphen 2-HDIV (58), and SIFT (59). We then plotted the ranked prediction of CADD, Polyphen-HDIV, and SIFT for the variants in *ITGA2B* and *ITGB3*. Variants were grouped as alloantigens, confirmed GT mutations, presumptive GT mutations, and novel variants. The ROC analysis was done through R package of pROC v1.7.3 ([cran.r-project.org/web/packages/pROC/index.html](http://cran.r-project.org/web/packages/pROC/index.html)) (90).

To examine the concordance between the different algorithms, the ranked scores from the different methods were first displayed as a heat map. We also displayed the ranked scores for all of the variants by group for each algorithm.

**Statistical Analysis.** Each experiment was repeated at least four times. Results are expressed as mean  $\pm$  SEM with number of observations  $n$ . Data were analyzed using GraphPad Prism 5 software. Significant differences were determined using Student  $t$  test and considered significant at  $P \leq 0.05$ .

The CI of ROC analysis is calculated by CI function of pROC package (with default parameters (95% CI through 2,000 stratified bootstrap replicates) (90).

**ACKNOWLEDGMENTS.** We thank Dr. Davide Provasi for technical assistance, and Drs. Paul Bray, Peter Newman, Jean-Laurent Casanova, and Yuval Itan for reviewing the manuscript and helpful suggestions. This study was supported, in part, by Grant HL19278 from the National Heart, Lung and Blood Institute, Grant UL1 TR000043 from the National Center for Advancing Translational Sciences, National Institutes of Health Clinical and Translational Science Award program, and funds from Stony Brook University.

- Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. *Cell* 155(1):27–38.
- Tennesen JA, et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337(6090):64–69.
- Nelson MR, et al. (2012) An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 337(6090):100–104.
- Coller BS, Shattil SJ (2008) The GPIIb/IIIa (integrin  $\alpha$ IIb $\beta$ 3) odyssey: A technology-driven saga of a receptor with twists, turns, and even a bend. *Blood* 112(8):3011–3025.
- Coller BS, Seligsohn U, Peretz H, Newman PJ (1994) Glanzmann thrombasthenia: New insights from an historical perspective. *Semin Hematol* 31(4):301–311.
- Nurden AT, Fiore M, Nurden P, Pillois X (2011) Glanzmann thrombasthenia: A review of *ITGA2B* and *ITGB3* defects with emphasis on variants, phenotypic variability, and mouse models. *Blood* 118(23):5996–6005.
- Caen JP, et al. (1966) Congenital bleeding disorders with long bleeding time and normal platelet count. I. Glanzmann's thrombasthenia (report of 15 patients). *Am J Med* 41(1):4–26.
- Nurden AT, Caen JP (1974) An abnormal platelet glycoprotein pattern in three cases of Glanzmann's thrombasthenia. *Br J Haematol* 28(2):253–260.
- Phillips DR, Agin PP (1977) Platelet membrane defects in Glanzmann's thrombasthenia. Evidence for decreased amounts of two major glycoproteins. *J Clin Invest* 60(3):535–545.
- French DL, Seligsohn U (2000) Platelet glycoprotein IIb/IIIa receptors and Glanzmann's thrombasthenia. *Arterioscler Thromb Vasc Biol* 20(3):607–610.
- Negri A, Li J, Naini S, Coller BS, Filizola M (2012) Structure-based virtual screening of small-molecule antagonists of platelet integrin  $\alpha$ IIb $\beta$ 3 that do not prime the receptor to bind ligand. *J Comput Aided Mol Des* 26(9):1005–1015.
- Provasi D, Negri A, Coller BS, Filizola M (2014) Talin-driven inside-out activation mechanism of platelet integrin  $\alpha$ IIb $\beta$ 3 probed by multimicrosecond, all-atom molecular dynamics simulations. *Proteins* 82(12):3231–3240.
- Zhu J, et al. (2012) Structure-guided design of a high-affinity platelet integrin  $\alpha$ IIb $\beta$ 3 receptor antagonist that disrupts  $Mg^{2+}$  binding to the MIDAS. *Sci Transl Med* 4(125):125ra132.
- Zhu J, et al. (2010) Closed headpiece of integrin  $\alpha$ IIb $\beta$ 3 and its complex with an  $\alpha$ IIb $\beta$ 3-specific antagonist that does not induce opening. *Blood* 116(23):5050–5059.



15. Murcia M, Jiroukova M, Li J, Collier BS, Filizola M (2008) Functional and computational studies of the ligand-associated metal binding site of beta3 integrins. *Proteins* 71(4):1779–1791.
16. Filizola M, Hassan SA, Artoni A, Collier BS, Weinstein H (2004) Mechanistic insights from a refined three-dimensional model of integrin alphaIIb beta3. *J Biol Chem* 279(23):24624–24630.
17. Choi WS, Rice WJ, Stokes DL, Collier BS (2013) Three-dimensional reconstruction of intact human integrin alphaIIb beta3: New implications for activation-dependent ligand binding. *Blood* 122(26):4165–4171.
18. Zhu J, et al. (2008) Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. *Mol Cell* 32(6):849–861.
19. Eng ET, Smaghe BJ, Walz T, Springer TA (2011) Intact alphaIIb beta3 integrin is extended after activation as measured by solution X-ray scattering and electron microscopy. *J Biol Chem* 286(40):35218–35226.
20. Xiao T, Takagi J, Collier BS, Wang JH, Springer TA (2004) Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature* 432(7013):59–67.
21. Provasi D, Murcia M, Collier BS, Filizola M (2009) Targeted molecular dynamics reveals overall common conformational changes upon hybrid domain swing-out in beta3 integrins. *Proteins* 77(2):477–489.
22. Rosenberg N, et al. (1997) Glanzmann thrombasthenia caused by an 11.2-kb deletion in the glycoprotein IIIa (beta3) is a second mutation in Iraqi Jews that stemmed from a distinct founder. *Blood* 89(10):3654–3662.
23. Rosenberg N, et al. (2005) A 13-bp deletion in alpha(IIb) gene is a founder mutation that predominates in Palestinian-Arab patients with Glanzmann thrombasthenia. *J Thromb Haemost* 3(12):2764–2772.
24. Fiore M, Pillois X, Nurden P, Nurden AT, Austerlitz F (2011) Founder effect and estimation of the age of the French Gypsy mutation associated with Glanzmann thrombasthenia in Manouche families. *Eur J Hum Genet* 19(9):981–987.
25. Seligsohn U, Rososhansky S (1984) A Glanzmann's thrombasthenia cluster among Iraqi Jews in Israel. *Thromb Haemost* 52(3):230–231.
26. Khanduri U, et al. (1981) Glanzmann's thrombasthenia. A review and report of 42 cases from South India. *Thromb Haemost* 46(4):717–721.
27. Schlegel N, et al. (1995) The molecular genetic basis of Glanzmann's thrombasthenia in a gypsy population in France: Identification of a new mutation on the alpha IIb gene. *Blood* 86(3):977–982.
28. Awidi AS (1992) Rare inherited bleeding disorders secondary to coagulation factors in Jordan: A nine-year study. *Acta Haematol* 88(1):11–13.
29. French DL, Collier BS (1997) Hematologically important mutations: Glanzmann thrombasthenia. *Blood Cells Mol Dis* 23(1):39–51.
30. Nurden AT, Pillois X, Nurden P (2012) Understanding the genetic basis of Glanzmann thrombasthenia: Implications for treatment. *Expert Rev Hematol* 5(5):487–503.
31. Peterson JA, McFarland JG, Curtis BR, Aster RH (2013) Neonatal alloimmune thrombocytopenia: Pathogenesis, diagnosis and management. *Br J Haematol* 161(1):3–14.
32. Kashiwagi H, et al. (2013) Demonstration of novel gain-of-function mutations of alphaIIb beta3: Association with macrothrombocytopenia and Glanzmann thrombasthenia-like phenotype. *Mol Genet Genomic Med* 1(2):77–86.
33. Kunishima S, et al. (2011) Heterozygous ITGA2B R995W mutation inducing constitutive activation of the alphaIIb beta3 receptor affects proplatelet formation and causes congenital macrothrombocytopenia. *Blood* 117(20):5479–5484.
34. Ghevaert C, et al. (2008) A nonsynonymous SNP in the ITGB3 gene disrupts the conserved membrane-proximal cytoplasmic salt bridge in the alphaIIb beta3 integrin and cosegregates dominantly with abnormal proplatelet formation and macrothrombocytopenia. *Blood* 111(7):3407–3414.
35. Nurden AT, Pillois X, Fiore M, Heilig R, Nurden P (2011) Glanzmann thrombasthenia-like syndromes associated with macrothrombocytopenias and mutations in the genes encoding the alphaIIb beta3 integrin. *Semin Thromb Hemost* 37(6):698–706.
36. Peyruchaud O, et al. (1998) R to Q amino acid substitution in the GFFKR sequence of the cytoplasmic domain of the integrin IIb subunit in a patient with a Glanzmann's thrombasthenia-like syndrome. *Blood* 92(11):4178–4187.
37. Schaffner-Reckinger E, et al. (2009) Overexpression of the partially activated alpha(IIb)beta3D723H integrin salt bridge mutant downregulates RhoA activity and induces microtubule-dependent proplatelet-like extensions in Chinese hamster ovary cells. *J Thromb Haemost* 7(7):1207–1217.
38. Jayo A, et al. (2010) L718P mutation in the membrane-proximal cytoplasmic tail of beta 3 promotes abnormal alpha IIb beta 3 clustering and lipid microdomain coalescence, and associates with a thrombasthenia-like phenotype. *Haematologica* 95(7):1158–1166.
39. Bledzka K, Smyth SS, Plow EF (2013) Integrin alphaIIb beta3: From discovery to efficacious therapeutic target. *Circ Res* 112(8):1189–1200.
40. Bosch X, Marrugat J, Sanchis J (2013) Platelet glycoprotein IIb/IIIa blockers during percutaneous coronary intervention and as the initial medical treatment of non-ST segment elevation acute coronary syndromes. *Cochrane Database Syst Rev* 11:CD002130.
41. Cox D, Brennan M, Moran N (2010) Integrins as therapeutic targets: Lessons and opportunities. *Nat Rev Drug Discov* 9(10):804–820.
42. Hynes RO (2002) Integrins: Bidirectional, allosteric signaling machines. *Cell* 110(6):673–687.
43. Avraamides CJ, Garmy-Susini B, Varner JA (2008) Integrins in angiogenesis and lymphangiogenesis. *Nat Rev Cancer* 8(8):604–617.
44. Bennett JS, Berger BW, Billings PC (2009) The structure and function of platelet integrins. *J Thromb Haemost* 7(Suppl 1):200–205.
45. Du X, Ginsberg MH (1997) Integrin alpha IIb beta 3 and platelet function. *Thromb Haemost* 78(1):96–100.
46. Mor-Cohen R, et al. (2012) Unique disulfide bonds in epidermal growth factor (EGF) domains of beta3 affect structure and function of alphaIIb beta3 and alphaV beta3 integrins in different manner. *J Biol Chem* 287(12):8879–8891.
47. Jallu V, Poulain P, Fuchs PF, Kaplan C, de Brevern AG (2012) Modeling and molecular dynamics of HPA-1a and -1b polymorphisms: Effects on the structure of the beta3 subunit of the alphaIIb beta3 integrin. *PLoS ONE* 7(11):e47304.
48. Mehrbod M, Trisno S, Mofrad MR (2013) On the activation of integrin alphaIIb beta3: Outside-in and inside-out pathways. *Biophys J* 105(6):1304–1315.
49. Laguerre M, et al. (2013) Molecular dynamics analysis of a novel beta3 Pro189Ser mutation in a patient with Glanzmann thrombasthenia differentially affecting alphaIIb beta3 and alphaV beta3 expression. *PLoS ONE* 8(11):e78683.
50. Blue R, et al. (2010) Effects of limiting extension at the alphaIIb beta3 on ligand binding to integrin alphaIIb beta3. *J Biol Chem* 285(23):17604–17613.
51. Vomund AN, Stuhlsatz-Krouper S, Dimitry J, Song Y, Frazier WA (2008) A naturally occurring extracellular alpha-beta clasp contributes to stabilization of beta3 integrins in a bent, resting conformation. *Biochemistry* 47(44):11616–11624.
52. Levin L, et al. (2013) A single disulfide bond disruption in the beta3 integrin subunit promotes thiol/disulfide exchange, a molecular dynamics study. *PLoS ONE* 8(3):e59175.
53. Kamata T, et al. (2010) Structural requirements for activation in alphaIIb beta3 integrin. *J Biol Chem* 285(49):38428–38437.
54. Valdar WS, Thornton JM (2001) Protein-protein interfaces: Analysis of amino acid conservation in homodimers. *Proteins* 42(1):108–124.
55. Valdar WS (2002) Scoring residue conservation. *Proteins* 48(2):227–241.
56. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
57. Kircher M, et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46(3):310–315.
58. Adzhubei IA, et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7(4):248–249.
59. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31(13):3812–3814.
60. Losonczy G, et al. (2007) Three novel mutations in the glycoprotein IIb gene in a patient with type II Glanzmann thrombasthenia. *Haematologica* 92(5):698–701.
61. Wright S (1937) The distribution of gene frequencies in populations. *Proc Natl Acad Sci USA* 23(6):307–320.
62. Basani RB, et al. (2000) A naturally occurring mutation near the amino terminus of alphaIIb defines a new region involved in ligand binding to alphaIIb beta3. *Blood* 95(1):180–188.
63. D'Andrea G, et al.; GLanzmann's Thrombasthenia Italian Team (GLATIT) (2002) Glanzmann's thrombasthenia: Identification of 19 new mutations in 30 patients. *Thromb Haemost* 87(6):1034–1042.
64. Nair S, Ghosh K, Shetty S, Mohanty D (2005) Mutations in GPIIb molecule as a cause for Glanzmann thrombasthenia in Indian patients. *J Thromb Haemost* 3(3):482–488.
65. Vinciguerra C, et al. (2001) Description of 10 new mutations in platelet glycoprotein IIb (alphaIIb) and glycoprotein IIIa (beta3) genes. *Platelets* 12(8):486–495.
66. Collier BS, et al. (1986) Immunologic and biochemical characterization of homozygous and heterozygous Glanzmann thrombasthenia in the Iraqi-Jewish and Arab populations of Israel: Comparison of techniques for carrier detection. *Br J Haematol* 62(4):723–735.
67. Keinan A, Clark AG (2012) Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 336(6082):740–743.
68. Coventry A, et al. (2010) Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat Commun* 1:131.
69. Schmitt MW, et al. (2012) Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci USA* 109(36):14508–14513.
70. Browning BL, Browning SR (2013) Detecting identity by descent and estimating genotype error rates in sequence data. *Am J Hum Genet* 93(5):840–851.
71. Schafer CM, et al. (2013) Whole exome sequencing reveals minimal differences between cell line and whole blood derived DNA. *Genomics* 102(4):270–277.
72. Collier BS, Folts JD, Smith SR, Scudder LE, Jordan R (1989) Abolition of in vivo platelet thrombus formation in primates with monoclonal antibodies to the platelet GPIIb/IIIa receptor. Correlation with bleeding time, platelet aggregation, and blockade of GPIIb/IIIa receptors. *Circulation* 80(6):1766–1774.
73. Wagner CL, et al. (1996) Analysis of GPIIb/IIIa receptor number by quantification of 7E3 binding to human platelets. *Blood* 88(3):907–914.
74. Reichert N, Seligsohn U, Ramot B (1975) Clinical and genetic aspects of Glanzmann's thrombasthenia in Israel: Report of 22 cases. *Thromb Diath Haemorrh* 34(3):806–820.
75. Cronberg S, Nilsson IM, Zetterqvist E (1967) Investigation of a family with members with both severe and mild degree of thrombasthenia. *Acta Paediatr Scand* 56(2):189–197.
76. Stormorken H, Gogstad GO, Solum NO, Pande H (1982) Diagnosis of heterozygotes in Glanzmann's thrombasthenia. *Thromb Haemost* 48(2):217–221.
77. Levin J (2013) The evolution of mammalian platelets. *Platelets*, ed Michelson AD (Academic, London), 3rd Ed, pp 3–26.
78. Schmaier AA, et al. (2011) Occlusive thrombi arise in mammals but not birds in response to arterial injury: Evolutionary insight into human cardiovascular disease. *Blood* 118(13):3661–3669.
79. Veltman JA, Brunner HG (2012) De novo mutations in human genetic disease. *Nat Rev Genet* 13(8):565–575.
80. Duquesnoy RJ (2012) The antibody response to an HLA mismatch: A model for non-self-self discrimination in relation to HLA epitope immunogenicity. *Int J Immunogenet* 39(1):1–9.

81. Carrasco Pro S, Zimic M, Nielsen M (2014) Improved pan-specific MHC class I peptide-binding predictions using a novel representation of the MHC-binding cleft environment. *Tissue Antigens* 83(2):94–100.
82. Kaplan CNH, Freedman J (2013) Alloimmune thrombocytopenia. *Platelets*, ed Michelson AD (Academic, London), 3rd Ed, pp 953–970.
83. Abecasis GR, et al.; 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.
84. Fu W, et al.; NHLBI Exome Sequencing Project (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493(7431):216–220.
85. Lau TL, Kim C, Ginsberg MH, Ulmer TS (2009) The structure of the integrin  $\alpha$ IIb $\beta$ 3 transmembrane complex explains integrin transmembrane signalling. *EMBO J* 28(9):1351–1361.
86. Yang J, et al. (2009) Structure of an integrin  $\alpha$ IIb $\beta$ 3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation. *Proc Natl Acad Sci USA* 106(42):17729–17734.
87. Metcalf DG, et al. (2010) NMR analysis of the  $\alpha$ IIb $\beta$ 3 cytoplasmic interaction suggests a mechanism for integrin regulation. *Proc Natl Acad Sci USA* 107(52):22481–22486.
88. Wang C, Bradley P, Baker D (2007) Protein-protein docking with backbone flexibility. *J Mol Biol* 373(2):503–519.
89. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8(3):275–282.
90. Robin X, et al. (2011) pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77.